APPENDIX E

EXPERT ELICITATION IN GEOLOGICAL AND GEOTECHNICAL ENGINEERING APPLICATIONS

E-1. Background. Many engineering evaluations are not amenable to quantitative analytical methods to determine the probability of unsatisfactory performance. Expert elicitation is one of several methods acceptable in USACE guidance documents for use in reliability analyses and risk assessments and has been used for various Corps projects. Expert elicitation is the formal quantification of expert opinion into judgmental probabilities. This document discusses how to structure a process of expert elicitation such that defensible probabilities result.

Risk analysis involves a large number of considerations - only a fraction of which are amenable to modeling and analysis. An analysis of modes of geotechnical failure (Wolff, 1998) indicated that expert elicitation was an appropriate probabilistic approach for analysis of seepage and piping through embankments, seepage through rock foundations, rock foundation stability, and erosion of soil and rock. The use of expert opinion in risk analysis allows the inclusion of uncertainties that might otherwise be difficult to calculate or quantify. Experienced engineers have long been required to evaluate opinions on many of these uncertainties. Judgmental probability is one way to quantitatively incorporate such evaluations into risk analysis.

The mathematical theory of probability is satisfied as long as the probabilities of exclusive and exhaustive events sum to 1.0. Thus, in some applications, probability is taken to mean the relative frequency of an event in a large number of trials; whereas, in others it is taken to mean the degree of belief that some event will occur or is true. Both interpretations are scientifically valid; judgmental probability is based on the latter.

On a basic level, judgmental probability is related to ones willingness to take action in the face of uncertainty. In practice, the magnitude of judgmental uncertainty can be compared to uncertainties in other situations, which may involve repetitive events, such as simple games of chance. If there is a greater willingness to bet on drawing a heart from a deck of cards than on the potential existence of piping within the foundation of a dam, then the judgmental probability of that adverse condition must be less than 1/4.

E-2. A Systematic Process to Elicit Quantified Judgmental Probabilities. The elicitation process needs to help experts think about uncertainty, to instruct and clarify common errors in how they quantify uncertainty, and to establish checks and balances to help improve the consistency with which probabilities are assessed. The process should not be approached as a ‘cookbook’ procedure; however, it is important that a systematic process be used to obtain defensible results. Based on experience at various Corps projects, it is recommended that the following steps be used when eliciting expert judgment. Details, of course, should be tailored to special needs; consequently, one or more of these steps may be eliminated.

a. Prepare background data and select issues. The initiator of the risk assessment should perform the following tasks in advance of the expert elicitation panel.

(1) Assemble and review all relevant site-specific data; visit the site; review related generic case histories.

(2) Develop and screen potential failure mechanisms for the site.

(3) Construct a preliminary event or fault tree for the site that includes all relevant modes of failure. Construct an additional event or fault tree for each potential remediation alternative for which judgmental probabilities are needed.

An event tree is a drawing that lays out the possible chains of events that might lead to adverse performance. The tree starts at the left with some initiating event, and then considers all possible chains of events that might lead from that first event (Figure E-1). Some of these chains lead to adverse outcomes; some do not. For each event in the tree, a conditional probability is assessed, presuming the occurrence of all the preceding events. The probability of a chain of events is obtained from the product of the probabilities of the events composing that chain.

A fault tree is a drawing that lays out possible sets of flaws in an engineered system that might lead to adverse performance. The tree starts at the right with some performance condition, and then considers the sets of faults (flaws) in the system that could have caused the adverse performance (Figure E-2). Most of these faults could only occur if earlier faults had occurred, and thus the tree is extended backward. Conditional probabilities for each fault are assessed as for an event tree, but the probability of a set of faults is calculated by starting at the adverse performance and moving backward.

Event trees are often easier for experts to conceive, but may become too complex. Fault trees, which focus only on adverse performance, may fail to uncover important combinations of events. Event and fault trees require a strict structuring of a problem into sequences. This allows probabilities to be decomposed into manageable pieces, and provides the accounting scheme by which those probabilities are put back together. In the process of decomposing a problem, it is sometimes helpful to construct an influence diagram that shows the inter-relationships of events, processes, and uncertainties. This diagram can be readily transformed into an event or fault tree.

Event and fault trees disaggregate failure sequences into the smallest pieces that can realistically be defined and analyzed, and can only be used for failure modes that are reasonably well understood. Failure modes, such as piping, for which the failure mechanism is poorly defined, cannot be further decomposed. Where the failure mechanism is well understood, it is usually good practice to disaggregate a problem such that component probabilities fall with the range [0.01 - 0.99], or better still, [0.1 - 0.9].

(4) Ensure a complete understanding of how the results of the expert elicitation will be

used by others involved in the risk assessment process.

(5) Select issues and uncertainties relative to the event trees that need to be assessed during expert elicitation. Form the issues into specific questions for the expert panel.

The questions should be carefully selected to represent the issues of concern and to achieve the desired objectives.

The initiator is now ready to use the expert elicitation process to formulate judgmental probabilities for all relevant modes of failure for each potential remediation alternative.

b. <u>Select a balanced panel of experts</u>. The choice of experts is the most important step in determining success or failure. Individuals selected must have an open mind and be willing to objectively judge different hypotheses and opinions that are not their own. Depending on personality and experience, experts may be individuals with special knowledge or individuals with a strongly argued point of view.

The panel must have a facilitator who is an individual versed in the issues who manages and encourages panel activities. The facilitator should be unbiased with respect to the outcome of the expert elicitation process and the facilitator must take care so that the expert panel's unbiased opinion is solicited, aggregated, and documented. Experts can be solicited both from within and from outside the initiator's organization. Appropriate USACE guidance on the use of technical experts should be followed. It is important that all panel experts be willing to be objective, commit time, and interact with others in a professional manner. The elicitation process has been shown to be most successful with between four and seven participants. A support team may be present to address panel questions regarding the site or specific events being investigated.

c. <u>Refine the issues with the panel, and decide on the specific uncertainties</u>. This phase sets up the problem, identifies specific uncertainties to be addressed, and defines the structure among those uncertainties. The goals are clear definitions of the uncertainties to be assessed, making unstated assumptions explicit, and dividing the technical problem into components with which experts can readily deal. For time-dependent requirements, it is best to request cumulative probabilities at different points in time from the panel. A minimum of three different dates should be requested so that a hazard function can be calculated form a data fit of the cumulative values provided by the panel.

A review package should be distributed to the experts well in advance. This package may include critical assumptions, interpretation of the foundation and other features, selected design parameters, analyses conducted, graphs and tables comparing principle issues, related generic case histories, proposed remedial measures, and the documentation for the construction of the event or fault tree. Time should be allocated at the initial meeting of the expert panel to review this information, ask clarifying questions, hold discussions with personnel knowledgeable about the site, and to visit the site if practicable.

All involved in the expert elicitation process must fully understand how the probability values and decisions made by the expert panel will be used and mathematically manipulated by other elements involved in the risk or reliability analysis. Economic accounting procedures may apply the numbers generated by the elicitation process and arrive at conclusions and recommendations far different than the experts envisioned.

d. Train the experts and eliminate error in eliciting judgmental probability. The training phase develops rapport with the experts, explains why and how judgmental probabilities are elicited, and how results will be used. Experts may be reluctant to participate unless assured about the intended use of the outcomes. During this phase, the philosophy of judgmental probability is reviewed, and an attempt is made to bring motivational biases out into the open.

The initiators of a risk assessment using expert elicitation must make every attempt to avoid the introduction of errors and bias into the result. Experts are known to display the following patterns when quantifying judgmental probability (Kahneman, Slovic, and Tversky, 1982).

(1) When asked to estimate the probability of an event, experts tend to assess larger values for the occurrence of those events that come readily to mind.

(2) When asked to estimate a numerical value, experts tend to fix on an initial estimate and then adjust for uncertainty by moving only slightly away from this first number.

(3) When asked to estimate the probability that an event ‘A’ originates from some process ‘B’, experts tend to base their estimate on the extent to which A resembles B rather than on statistical reasoning.

(4) An expert who perceives that he has had control over the collection or analysis of data tends to assign more credibility to the results than does an expert who only reviews the results.

The above patterns may lead to several errors including overconfidence, insensitivity to base rate probabilities, insensitivity to sample size, misconceptions of chance and neglect of regression effects.

The simplest manifestation of overconfidence occurs when people are asked to estimate the numerical value of some unknown quantity, and then to assess probability bounds on that estimate. For example, a person might be asked to estimate the undrained shear strength of a foundation clay, and then asked to assess the 10 and 90 percent bounds on that estimate.

People typically respond with probability bounds that are much narrower than empirical results suggest they should be.

Another display of overconfidence occurs when people are asked to estimate the numerical value of either small (<0.1) or large (>0.9) probabilities. People consistently underestimate low probabilities (unusually low shear strength) and overestimate high probabilities (continued satisfactory performance of a structure). Empirical results verify this effect (Lichenstein, Fischoff, and Phillips, 1982). With training people can learn to calibrate their estimates of probabilities between 0.1 and 0.9 (Winkler and Murphy, 1977). However, when required to estimate probabilities outside this interval, people error due to overconfidence. Research also suggests that the harder the estimation task the greater the overconfidence.

The simplest manifestation of insensitivity to base-rate occurs when people focus on recent information while ignoring background rates. For example, regional rates of seismic events provide important information about risk; yet this background information may be discounted if site reconnaissance fails to uncover direct evidence of a seismic hazard – even though the reconnaissance may be geologically inconclusive.

Insensitivity to sample size occurs when people presume that the attributes (averages, standard deviations) of small samples are close to the attributes of the populations from which the samples were taken. People tend to overemphasize the results of a small suite of samples even though the fluctuations in the attributes from one small suite to the next can be great.

Misconceptions of chance are familiar in the “gambler’s fallacy” that events average out. People expect that the essential attributes of a globally random process will be reflected locally. Local variations of soil properties about some spatial average are not corrected as more measurements are taken; they are just diluted with more data.

Neglect of regression effects occurs when people overlook the fact that in predicting one variable from another (e.g. dry density from compactive effort), the dependent variable will deviate less from its mean than will the independent variable. Exceptionally high compactive effort produces, on average, high – but not exceptionally high – densities; and the converse for exceptionally low compactive effort. Representativeness leads people to erroneously overlook this regression toward the mean.

Beyond these statistical errors, an additional source for error is motivational biases. These are factors, conscious or not, that lead to inaccurate or incomplete assessments. The desire to appear knowledgeable, and thus under report uncertainty or the desire to advance a special cause, and thus refuse to credit alternate points of view are typical examples.

The training phase explains the how people can quantify judgmental uncertainties, and how well judgmental probabilities compare to the real world. The goal is to encourage the experts to think critically about how they quantify judgment, and to avoid common sources of statistical errors and biases discussed above. The training phase might involve having experts explain how they think about uncertainty and how they use data in modifying uncertainties. A few warm-up exercises can expose systematic biases in the experts’ responses. “Thought experiments,” which have experts explain retrospectively how unanticipated outcomes of an engineering project might have occurred, serve to open up the range of considerations experts entertain.

e. Elicit the judgmental probabilities of individual experts in quantified degrees of belief. This phase develops numerical probabilities for the component events or faults identified in the structuring phase. The goal is to obtain coherent, well-calibrated numerical representations of judgmental probability for individual experts on the panel, and to aggregate these into the probabilities for the entire panel. This is accomplished by presenting comparative assessments of uncertainty to the panel members and interactively working toward probability distributions.

(1) Associate probabilities with descriptive statements. In the early phases of expert elicitation, people find verbal descriptions more intuitive than they do numbers. Such descriptions are sought for the branches of an event or fault tree. Empirical translations are then used to approximate probabilities (Table E-1). This technique has been shown to improve consistency in estimating probabilities among experts. However, the range of responses is large, and the probabilities that an expert associates with verbal descriptions often changes with context.

(2) Avoid intuitive or direct assignment of probabilities. It is common for experts who have become comfortable using verbal descriptions to wish to directly assign numerical values to those probabilities. This should be discouraged, at least initially. The opportunity for systematic error or bias in directly assigning numerical probabilities is great. More experience with the process on the part of the experts should be allowed to occur before directly assigning numbers. At this initial point, no more than order of magnitude bounds on the elicited numerical degrees of belief is a realistic goal.

Table E-1. Empirical Translations of Verbal Descriptions of Uncertainty

Verbal Description	Probability Equivalent	Low	High
Virtually impossible	0.01	0.00	0.05
Very unlikely	0.10	0.02	0.15
Unlikely	0.15	0.04	0.45
Fairly unlikely	0.25	0.02	0.75
Fair chance, even chance	0.50	0.25	0.85
Usually, likely	0.75	0.25	0.95
Probable	0.80	.030	0.99
Very probably	0.90	0.75	0.99
Virtually certain	0.99	0.90	1.00

Source: Vick (1997), and Lichtenstein and Newman (1967).

(3) Quantify probabilities of discrete events. The theory of judgmental probability is based on the concept that numerical probabilities are not intuitive. This means that the most accurate judgmental probabilities are obtained by having an expert compare the uncertainty of the discrete event in question with other, standard uncertainties as if he were faced with placing a bet.

For a practical example, consider a dam site at which a fault zone in the foundation is suspected. Some exploration has been carried out, but the results are not definitive. If the expert would prefer to bet on the toss of a coin rather than on the existence of the fault, the judgmental probability of the fault existing must be less than half (0.5). Should he prefer to bet on the existence of the fault over the roll of a six-sided die, then the judgmental probability of the fault existing would be greater than one-sixth (0.17), and so forth. Changing the payoff odds on the gambles is another way of bounding the assessment.

Research on expert elicitation has addressed a number of issues regarding whether questions should be expressed in terms of probabilities, percentages, odds ratios, or log-odds ratios. In dealing with relatively probable events, probabilities or percentages are often intuitively familiar to experts. However, with rare events, odds ratios (such as, “100 to 1”) may be easier because they avoid very small numbers. Definitive results for the use of aids such as probability wheels are lacking; and in the end, facilitators and experts must choose a protocol that is comfortable to the individuals involved.

(4) Quantify probability distributions. Not all uncertain quantities involve simple probabilities of discrete events. Many are defined over a scale, and the issue is to assess a judgmental probability distribution over that scale. For example, the friction between a concrete mass and its rock foundation should have a value between 0 and 90^o. A probability distribution summarizes the relative uncertainty about the parameter’s value lying within specific intervals of the scale. In expert elicitation it is convenient to represent probability distributions as cumulative functions, which graph the scale of the parameter along the horizontal axis, and the judgmental probability that the realized value is less than specific values along the vertical axis.

The process starts by asking the experts to suggest extreme values for the uncertainty quantity. It is useful to have the expert describe ways that values outside these extremes might occur. Then, the experts are asked to assess probabilities that values outside the extremes occur. Starting with extreme values rather than best estimates is important in guarding against overconfidence and anchoring. Asking the experts to conceive extreme scenarios makes those scenarios ‘available,’ and allows one to think about the extremes more readily.

As numerical values are elicited, the facilitator should begin plotting these on graph paper; however, at this point the plot should not be shown to the experts, because it may bias future responses to conform to the previous ones. As ever more assessments are made, they are plotted on the graph to begin establishing bounds and to point out inconsistencies.

In checking for consistency, it is useful to compare numerical results elicited as values to those elicited as probabilities. In the *fixed probability* approach, the expert is given a probability, and asked for a corresponding value of the uncertain quantity; or given a probability interval, and asked for corresponding ranges of the uncertain quantity. For example, “What value of the friction angle do you think has a 1/3 chance of being exceeded?” “What values of the friction angle do you think have a 50:50 chance of bounding the true value?” In the *fixed value* approach, the expert is given a value of the uncertain quantity and asked the probability that the true value is less than that value, or the expert is given a range of values and asked the probability that the true value lies within that range. For example, “Would you be more inclined to bet on the chance of the friction angle being within the range 25 to 35 degrees or on drawing a diamond from this deck of cards?” Limited research suggests that fixed value procedures produce probability distributions that are more diffuse and better calibrated than do fixed probability or interval procedures.

(5) Use of normalized or base-rate frequency to estimate probabilities. The normalized frequency approach for assessing judgmental probabilities starts with an observed, empirical frequency of similar events in dam inventories and allows the experts to adjust the rates to reflect local conditions. The approach is appealing in that it begins with empirical frequencies. On the other hand, the procedure increases anchoring bias and a number of issues make the procedure difficult to use in practice. These include, identifying a relevant subcategory of events in the dam inventories with which to compare the present project, the fact that dam incidents are seldom simple cause and effect, and the complex procedures and calculations involved in adjusting base-rate frequencies. This method should be used only with caution.

(6) Use of reliability analysis to assess probabilities. For some component events, engineering models are available for predicting behavior. In these cases, reliability analysis can be used to assess probabilities associated with the components. Reliability analysis propagates uncertainty in input parameters to uncertainties in predictions of performance. The assessment problem is changed from having experts estimating probabilities of adverse performance to estimating probability distributions for input parameters. Once probabilities for the input parameters are assessed, a variety of mathematical techniques can be used to calculate probabilities associated with performance. Among these are, first-order second-moment approximations, advance second-moment techniques, point-estimate calculations, or Monte Carlo simulation. Sometimes, experts elect to assess an additional component of uncertainty in the reliability analysis to account for model error. While there are many ways to do this, the most common is to assign a simple, unit-mean multiplier to the model output, having a standard deviation estimated by the experts to reflect model uncertainty.

f. Revise and combine individual probabilities into a consensus. Once the judgmental probabilities of individuals have been elicited, attention turns to aggregating those probabilities into a consensus of the panel. Consensus distributions often outperform individual experts in forecasting because errors average out (Rowe 1992). Both mathematical and behavioral procedures can be used to form consensus distributions.

After the initial round of elicitation, the value assessments may be plotted on a graph to establish bounds and to point out inconsistencies to the expert panel. The facilitator may wish to discuss “outlier” values with the individual expert(s) to ensure that the questions were understood. The experts should be given the opportunity to revise their opinions. Additional rounds of elicitation may be required until the panel is satisfied with the results.

Mathematical procedures typically use some form of weighted sum or average to aggregate individual probabilities. The weights, if not taken equal, are based on experts' self-weightings, on peer-weights, or on third-party weightings (e.g., by the 'evaluator'). Caution must be exercised if the experts group into "schools of thought," and thus do not give statistically independent answers (Ferrell, 1985).

Behavioral procedures involve an unstructured process in which experts discuss issues among themselves in order to arrive at a consensus judgment. The concept is straightforward. The potential information available to a group is at least as great as the sum of the information held by the individuals. It is presumed that errors are unmasked, and that the group discussion resolves ambiguities and conflict. Empirical evidence supports this contention, but strong involvement of a facilitator is the key to success of the expert elicitation process.

E-3. <u>Verifying and Documenting Judgmental Probabilities</u>. Once a set of probabilities has been elicited, it is important to ensure that the numerical probabilities obtained are consistent with probability theory. This can be done by making sure that simple things are true, such as the probabilities of complementary events adding up to 1.0. It is also good practice to restructure questions in logically equivalent ways to see if the answers change, or to ask redundant questions of the expert panel. The implications of the elicited probabilities for risk estimates and for the ordering of one set of risks against other sets is also useful feedback to the experts.

For credibility and defensibility, the process and results of an expert elicitation should be well documented, reproducible, subject to peer review, and neutral. The results of the process should also pass a “reality check” by the initiator’s organization. The process should be documented such that it is possible, in principal, to reproduce all the calculations involved and to arrive at the same answers. Calculation models should be fully specified. All questions asked and the responses of the experts should be tabulated. The source of all data and estimates in the study should be traceable to a person or a report. This means that the names of the expert panel members should be listed and the responses associated with each expert should be explicit.

E-4. <u>References</u>.

Ferrell, W., "Combining individuals judgments," in Wright, G. (Ed.), *Behavioral Decision Making*, Plenum, NY, 1985.

Kahneman, D., Slovic, P., and Tversky, A., (Eds.), *Judgment under uncertainly, heuristics and biases*, Cambridge University Press, Cambridge, 1982.

Lichtenstein, S., and Newman, J., "Empirical scaling of common verbal phrases associated with numerical probabilities," *Psychon. Sci., 9*(10): 563-564, 1967.

Lichtenstein, S., Fischhoff, G., and Phillips, L., "Calibration of Probabilities: The State of the Art to 1980," in Kahneman, D., Slovic, P., and Tversky, A., (Eds.), *Judgment under uncertainly, heuristics and biases*, Cambridge University Press, Cambridge, pp. 306-334, 1982..

Moser, D., "Risk Analysis Framework: Event Trees and Fault Trees," U.S. Army Corps of Engineers Seminar Notes, July 1997.

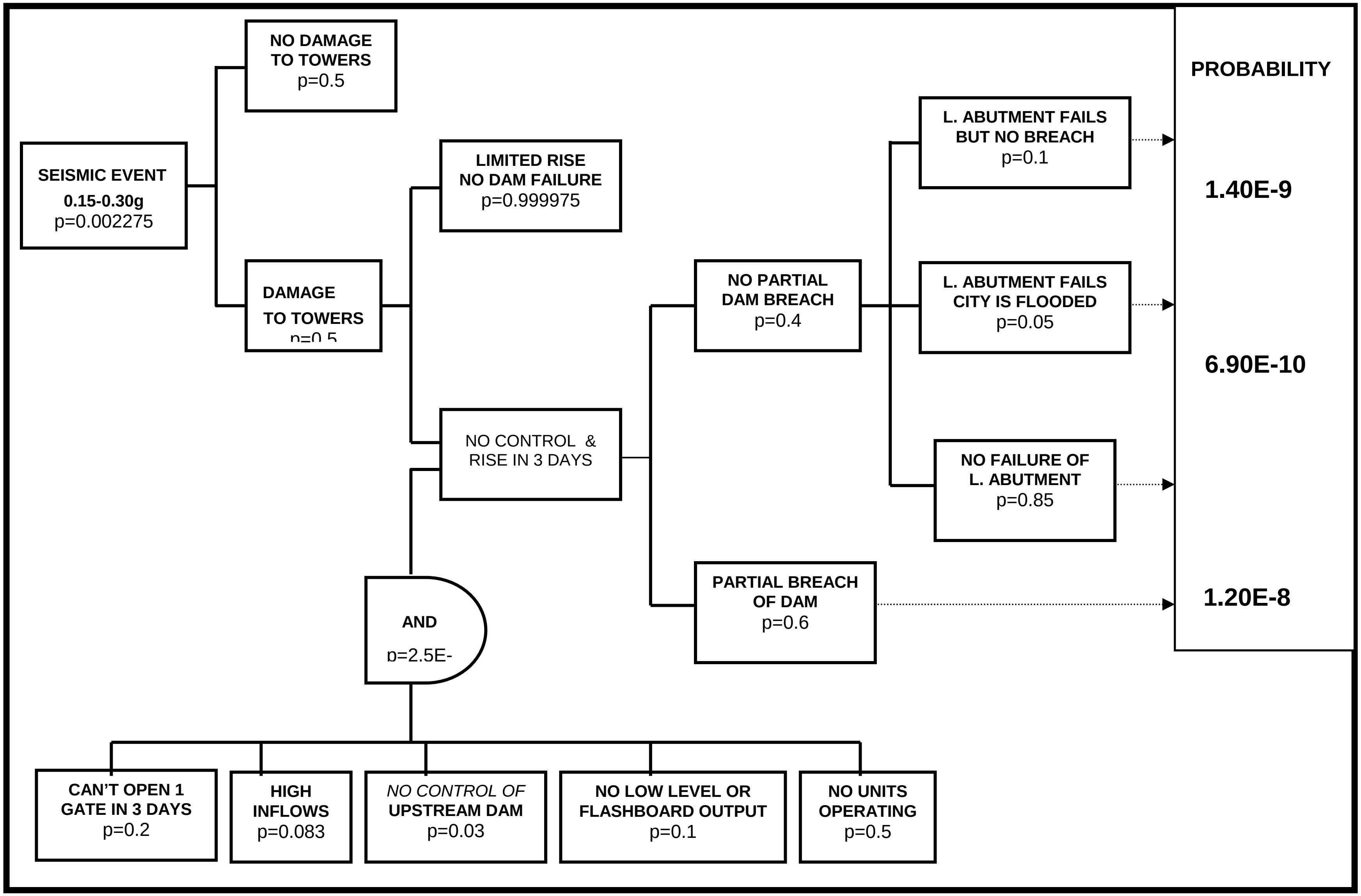
Rowe, G., "Perspectives on expertise in aggregation of judgments," in Wright, G., and Bolger, F. (Eds.), Expertise and Decision Support, Plenum Press, NY, pp. 155-180, 1992.

Salmon, G.M., "Risk Assessment for Dams – Better Tools Are Needed," in Hoeg, K. and Torblaa, E. (Eds.), *Risk-Based Dam Safety Evaluations*, Trondheim, 1997.

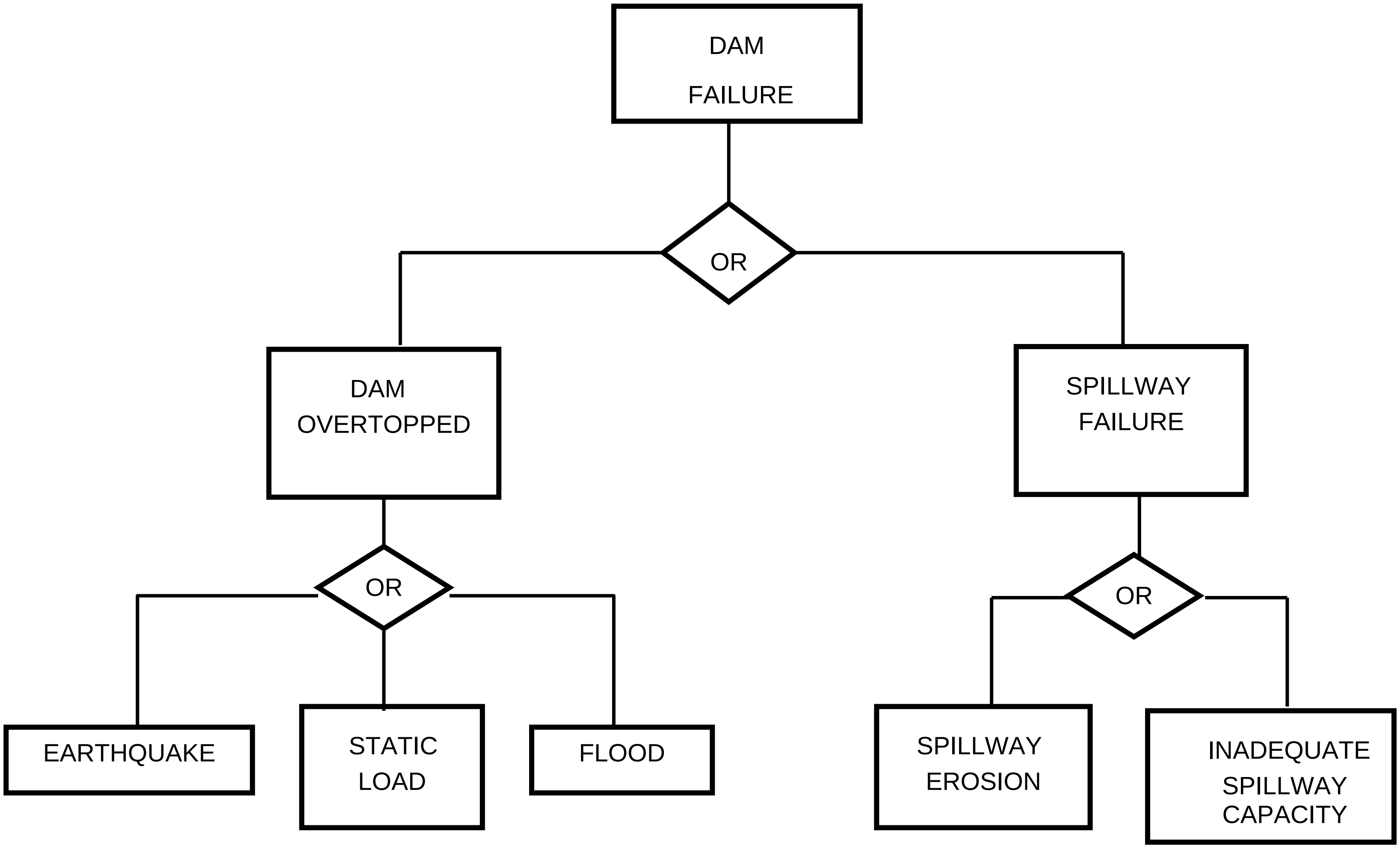
Vick, S, "Risk Analysis for Norwegian Dams" in Mitchell, J and Duncan, J.M. (Eds.), *Use of Probability in Geotechnical Aspects of Dams and Waterways*, Virginia Polytechnic, 1997.

Winkler, A., and Murphy, R., "Reliability of Subjective Probability Forecasts of Precipitation and Temperature," *Applied Statistics*, v. 26, 1977.

Wolff, T., "Assessment of Geotechnical Reliability", report to USACE, St. Louis District, 1998.



Event Tree: Earthquake loading of a dam.
Figure E-1.



FAULT TREE: Dam failure by overtopping or spillway failure.
Figure E-2

(D. MOSER)